Decoding the causal drivers of spatial cellular topology

Prannav Shankar[1], Huan Liang[1], Uthsav Chitra[*2], and Rohit Singh[*1]

[1]Department of Bioinformatics and Biostatistics, Duke University, Durham, NC, USA
[2]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Abstract

Decoding how cells influence and communicate with each other in space is fundamental for understanding tissue organization. However, existing approaches either overlook spatial context entirely or rely solely on local cell-cell adjacency, failing to capture how global tissue topology shapes cellular communication. Here, we present GLACIER, which introduces *spatial Granger causality* to infer transcriptional and signaling relationships that emerge from tissue organization. By combining GASTON's global isodepth coordinate with Velorama's graph-based causal inference framework, we enable bidirectional inference of regulatory relationships along spatial axes, identifying transcription factor-target interactions and ligand-receptor pairs that operate across spatial domains. Applying GLACIER to single-cell spatial transcriptomics data from the mouse cerebellum, we identify both continuous within-cell-type regulatory gradients and discontinuous drivers at layer interfaces, while distinguishing between forward and backward cellular communication along the isodepth axis. Our approach reveals how tissue architecture directs patterns of cellular communication, providing a framework for understanding spatially-encoded regulatory programs.

Software availability: `https://github.com/rohitsinghlab/glacier`.

[*]Correspondence to uchitra@broadinstitute.org and rohit.singh@duke.edu.

1 Introduction

Multicellular organisms depend on precisely orchestrated cell-cell interactions that unfold across multiple spatial scales, including both direct, short-range interactions and indirect, longer-range communication. For example, cells directly communicate with one another by releasing ligand molecules that bind to receptors on neighboring cells, triggering immediate signaling changes. More broadly, transcription factors (TFs) can initiate regulatory programs that cascade across cellular neighborhoods and lead to coordinated changes in gene expression programs in larger domains of a tissue [1]. Understanding these spatially-encoded cellular interactions is crucial for decoding many fundamental biological processes including tissue development [2], immune response [3], wound healing [4], and cancer progression [5].

Recent advances in spatial transcriptomics (ST) offer unprecedented potential for characterizing cellular interactions in tissues. Unlike traditional single-cell RNA sequencing (scRNA-seq) which dissociates cells before measurement and loses spatial information, ST technologies measure both gene expression *and* the spatial location of individual cells in tissue slices. However, there are two key challenges in the inference of cellular interactions from current ST datasets. First, transcript counts in ST data are highly sparse, making it difficult to reliably estimate short-range, *local* correlations in gene expression [6, 7]. For example, whole-transcriptome technologies such as Slide-SeqV2 [8] and Stereo-seq [9] measure a median of less than 500 unique molecular identifiers (UMIs) per cell. Second, and more fundamentally, understanding how cells communicate over long scales requires knowledge of the *global* geometry of the tissue and the spatial axis along which regulatory signals propagate.

Current computational methods for inferring cellular or regulatory interactions from ST data do not leverage global tissue organization. Nearly all methods implicitly assume a uniform interaction likelihood between neighboring cells by representing these local relationships through cell-cell adjacency graphs, and they look for correlated TF-gene or ligand-receptor expression along edges of such an adjacency graph [6, 10, 11, 12, 13, 7]. However, this oversimplified view fails to capture how intercellular signals propagate in specific directions determined by tissue architecture. For instance, endothelial cells communicate along paths defined by tissue cytoskeleton [14], while the Notch signaling pathway operates bidirectionally along particular spatial axes during development [15]. Accurate inference of cellular interactions thus requires incorporating global tissue structure to understand the directionality of gene expression variation.

GASTON [16] addresses this need by learning a "topographic map" of tissue from ST data. This map is defined by a spatial coordinate called "isodepth" which varies globally across the tissue slice, and measures the direction of maximum variation in gene expression. It allows inference of directional cellular interactions by structuring the tissue into a hierarchical framework. This approach helps distinguish between local interactions and broader regulatory signals dictated by global tissue topology. In the mouse cerebellum, GASTON successfully revealed both discrete cellular domains (layers) and continuous gradients of gene activity, providing the first comprehensive map of spatial gene expression in this tissue. However, GASTON models each the variation of gene independently across the isodepth coordinate and does not account for gene-gene interactions, leaving open the challenge of inferring cellular communication and regulatory relationships from ST data.

We introduce *spatial Granger causality*, a fundamentally new approach for learning cellular interactions from ST data. The key insight is that tissue organization, as captured by GASTON's isodepth, creates natural directions of information flow that can be represented as a directed acyclic graph (DAG). While classical Granger causality required data to follow a linear time series, recent advances [17, 18, 19] have enabled Granger causal inference on partial orderings represented by DAGs. Building on this foundation, we present GLACIER (Granger-Led Analysis of Cellular Isodepth and Expression Regulation), which combines GASTON's global spatial coordinate with Velorama's DAG-based nonlinear Granger causality to identify TF-gene and ligand-receptor relationships that propagate along spatial axes. GLACIER systematically captures how regulatory information flows through tissue structure, enabling directional inference of cellular

signaling interactions. Applied to Slide-SeqV2 data of mouse cerebellum [8, 20], GLACIER identifies directional transcriptional programs within layers, revealing that oligodendrocytes predominantly regulate neuronal connectivity in one direction. At the boundary between the granular and Purkinje-Bergmann layers, GLACIER reports regulatory activity that is distinct from what is seen in either layer individually, suggesting that inter-layer communication operates through specialized mechanisms rather than a simple gradient. These findings demonstrate how integrating spatial organization with causal inference can reveal fundamental principles of tissue architecture and cellular communication.

2 Methods

2.1 Granger causality in time-series

Granger causality is a statistical framework for inferring causal relationships in time-series data by assessing whether past values of one variable $\mathbf{x}$ (predictor) improve the prediction of another variable $\mathbf{y}$ (target) beyond what can be explained by the latter's own past values. The key principle is that a cause must precede its effect, and predictive relationships between variables may therefore indicate potential causality. Granger causal interactions are statistical estimates, best interpreted as causal hypotheses to be prioritized for experimental validation; for example, a Granger causal relationship between $\mathbf{x}$ and $\mathbf{y}$ may be indirect, mediated through a latent and potentially unobserved variable.

Linear Granger causality. Given two time series of length N, $\mathbf{x} = (x_t)_{t=1}^N \in \mathbb{R}^N$ and $\mathbf{y} = (y_t)_{t=1}^N \in \mathbb{R}^N$, we say that they follow a *vector autoregressive (VAR)* model with $L > 0$ lags if

$$y_t = \sum_{\ell=1}^{L} \alpha_\ell y_{t-\ell} + \sum_{\ell=1}^{L} \beta_\ell x_{t-\ell} + \epsilon_t, \tag{1}$$

where ϵ_t is a noise term, and $\alpha_\ell, \beta_\ell \in \mathbb{R}$ are coefficients that describe the contribution of the observations $y_{t-\ell}, x_{t-\ell}$, respectively, at the ℓ-th *lag*, i.e. the $(t-\ell)$-th time. The value $L > 0$ indicates how far back in time past observations influence the observation y_t, i.e. the *maximum lag*. VAR models are widely used for time-series analysis in finance and economics [21].

We say that $\mathbf{x}$ linearly *Granger causes* $\mathbf{y}$ if $\beta_\ell \neq 0$ for some lag $0 < \ell \leq L$ [22]. The standard approach for determining linear Granger causality is through hypothesis testing, where the null hypothesis is $H_0 : \beta_\ell = 0$ for all lags $\ell = 1, \ldots, L$, e.g. using an F-test or likelihood ratio test [23, 24].

Non-linear Granger causality. In many scenarios, the time series $\mathbf{x}, \mathbf{y}$ are not linearly related. Thus, we generalize (1) and say that $\mathbf{x}$ and $\mathbf{y}$ follow a *non-linear vector autoregressive (NVAR)* model with L lags if

$$y_t = f(y_{t-1}, \ldots, y_{t-L}) + g(x_{t-1}, \ldots, x_{t-L}) + \epsilon_t, \tag{2}$$

where $f, g : \mathbb{R}^L \to \mathbb{R}$ are functions— potentially parameterized by neural networks— that specify how the observation y_t at time t depends on the previous L observations $y_{t-1}, \ldots, y_{t-L}$ and $x_{t-1}, \ldots, x_{t-L}$, respectively, and ϵ_t is a noise term.

In a non-linear setting, Granger causality can be detected through two approaches: *ablation-* or *invariance-based* [17]. The approporiate choice depends on the problem setting and the kind of interactions modeled between predictor variables:

- **Ablation-based approach.** Causality is assessed by training two models: a full model $\mathcal{M}_{\text{full}}$ that includes past values of both $\mathbf{x}$ and $\mathbf{y}$, and a restricted model $\mathcal{M}_{\text{reduced}}$ that omits $\mathbf{x}$. If removing $\mathbf{x}$ significantly reduces predictive accuracy, $\mathbf{x}$ is deemed Granger-causal of $\mathbf{y}$.

- **Invariance-based approach.** Instead of explicitly comparing models, this approach applies regularization to $\mathbf{x}$-dependent terms. If removing $\mathbf{x}$ does not alter the model's predictions (e.g., its coefficients shrink to zero under sparsity constraints), then $\mathbf{x}$ does not Granger-cause $\mathbf{y}$. This method is computationally efficient and allows joint learning of multiple interactions. We note that applying it in a neural network setting requires special architecture and training considerations to encode the required sparsity constraints [17].

2.2 Spatial Granger causality

Classical Granger causality approaches require a total ordering on the N observations, and are not immediately applicable to *spatially structured* data where there is no obvious total ordering of locations in space.

Here we leverage recent extensions of Granger causality to *partially ordered* data described by a directed acyclic graph (DAG) [17, 18, 19]. To extend Granger causality to spatially structured data, we combine the DAG-structured Granger causality framework with a novel construction of a *spatial DAG*.

We start by describing the DAG-structured Granger causality framework of GrID-Net and Velorama [17, 18, 19]. Given a DAG $G = (V, E)$ and observations $\mathbf{x} = (x_v)_{v \in V} \in \mathbb{R}^{|V|}, \mathbf{y} = (y_v)_{v \in V} \in \mathbb{R}^{|V|}$ defined on each vertex $v \in V$, we say that the observations $\mathbf{x}, \mathbf{y}$ follow a *DAG-structured vector autoregressive (DVAR)* model with L lags, modifying (2) as follows:

$$y_v = f(\{y_u\}_{u \in \mathcal{P}_L(v)}) + g(\{x_u\}_{u \in \mathcal{P}_L(v)}) + \epsilon_v, \tag{3}$$

where $\mathcal{P}_L(v)$ is the set of vertices $u \in V$ that are ancestors of v and are at most distance L from v. That is,

$$\mathcal{P}_L(v) = \{u \in V \mid u \text{ is an ancestor of } v \text{ with } d_G(u, v) \leq L\}, \tag{4}$$

where $d_G(u, v)$ is the length of the shortest directed path from u to v in G; and $f, g : 2^{\mathbb{R}} \to \mathbb{R}$ are functions that map each subset of the observations $\mathbf{y}, \mathbf{x}$, respectively, to a real number. As mentioned before, one can then infer causal relationships using an ablation or invariance-based approach. To estimate Granger causal relationships between individual genes and ATAC-seq peaks, GrID-Net used an ablation based approach [18, 25]. In contrast, when seeking to assess and select from multiple TFs (i.e., predictors) acting combinatorially, Velorama used an invariance-based approach with per-TF sparsity constraints [17].

DAG construction. We construct a spatial DAG using a *spatial potential* function, similar to how [17] constructs a DAG from a pseudotime function. Briefly, suppose we are given N spatial locations $\mathbf{s}_1 = (x_1, y_1), \ldots, \mathbf{s}_N = (x_N, y_N)$ and a real number $d_i \in \mathbb{R}^i$ defined on each spatial location $\mathbf{s}_i$, which we call a *spatial potential*. We form a DAG $G = (V, E)$ whose vertices $V = \{s_1, \ldots, s_N\}$ are the spatial locations, and whose edges E are those in the k-nearest-neighbors (k-NN) graph. Each edge (i, j) is oriented in the direction of increasing spatial coordinate; that is for any two arbitrary cells i, j, we include a directed edge $i \to j$ if $d_i > d_j$. Such a construction guarantees that G is a DAG [17].

Isodepth. While our approach is generalizable to any spatial potential d_i, in practice we use the spatial potential d_i learned by GASTON [16]. The GASTON spatial potential d is called the *isodepth* and is a 1-D latent coordinate describing the direction of maximum spatial variation in gene expression. For example, in layered tissues (e.g. the cerebellum tissue slice that we analyze in Section 3 and Figures 2,4), the isodepth d corresponds to the *layer depth*. See [16] for more details. We note that isodepth is just an example spatial potential and our method can be applied directly to perform causal analysis in the setting of other spatial potentials (e.g., radial distance from a blood vessel, germinal center, or tumor mass).

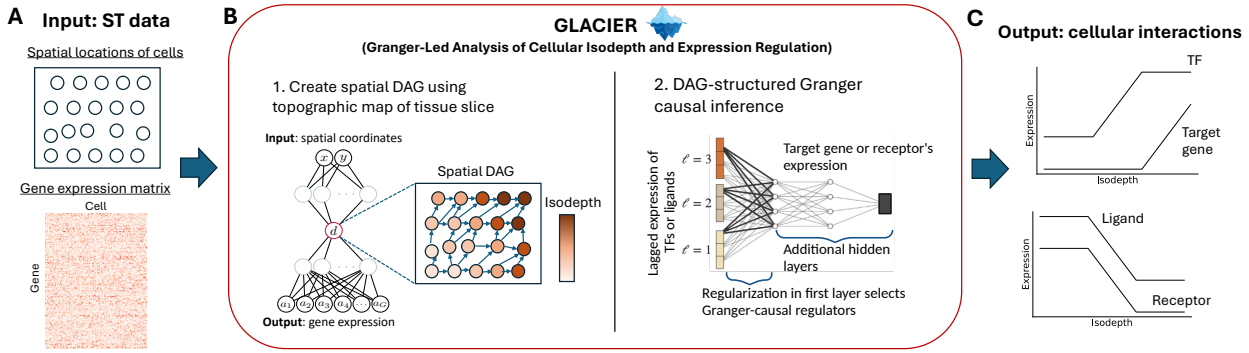

Figure 1: **(A)** The input to GLACIER is spatial transcriptomics (ST) data from a 2-D tissue slice, consisting of the spatial locations of measured cells and the gene expression matrix. **(B)** GLACIER (Granger-Led Analysis of Cellular Isodepth and Expression Regulation) first learns a topographic map of the tissue slice defined by an isodepth coordinate, and uses the topographic map to form a spatial DAG. GLACIER then performs directed acyclic graph (DAG)-structured Granger causal inference. **(C)** The output of GLACIER are cellular interactions: transcription factors (TFs) that are Granger causal for target genes, and ligand genes that are Granger causal for receptor genes.

Inference. Since our inference objectives (i.e. identifying combinations of genes as regulators) are similar to Velorama, we follow its invariance-based approach to extend Granger causality to spatial transcriptomics. For each candidate regulator $\mathbf{x} = (x_v)_{v \in V}$ (i.e. gene, out of N_G total candidates) and vertex $v \in V$ (i.e. cell) in G, we accumulate its expression across the $\mathcal{P}_L(v)$ ancestors of v. Together, these form the input to a neural network that predicts the expression of a target gene $\mathbf{y} = (y_v)_{v \in V}$. To enforce sparsity in the selection of regulators, we incorporate a regularization on the first layer of the neural network (Figure 1B,right). This acts as a feature selection mechanism, ensuring that only a subset of candidate regulators contribute to predicting the target gene. Specifically, the first hidden layer takes the form

$$h^{(1)} = \sigma\left(\sum_{\ell=1}^{L} (A')^{\ell} X W_1^{\ell} + b_1\right), \tag{5}$$

where $(A')^{\ell}$ is the ℓ-th power of the adjacency matrix of G; $X = [\mathbf{x}] \in \mathbb{R}^{|V| \times N_G}$ is the regulator expression matrix, where each column is a candidate regulator $\mathbf{x}$; $W_1^{\ell} \in \mathbb{R}^{N_G \times d}$ are trainable weight matrices with a lasso penalty; d is the dimensionality of the hidden layers; and $\sigma(\cdot)$ is a non-linear activation function.

Following [17], we apply a hierarchical regularization scheme that penalizes W_1^{ℓ}. Longer lags are more heavily penalized than shorter ones. This encourages robustness in selecting lagged interactions while reducing sensitivity to the choice of the maximum lag L [26]. To ensure sparsity, the model is trained using proximal gradient descent, which is well-suited for enforcing sparsity constraints. The regularization weight is specified by the hyperparameter λ. To select an appropriate λ, we perform a grid search over a range of values and ensemble the results across valid λ settings where some, but not all, weights remain nonzero. This enables GLACIER to infer spatially structured regulatory relationships that capture both local and long-range dependencies dictated by tissue architecture. We say that the *lag* of an interaction is the maximum number of non-zero weights. See [17] for details.

2.3 GLACIER algorithm and implementation

We present GLACIER, an algorithm for identifying Granger causal interactions from single-cell, spatial transcriptomics (ST) data (Figure 1A).

Input. The input to GASTON is ST data $(\mathbf{A}, \mathbf{S})$ which consists of

- an $N \times G$ gene expression matrix $\mathbf{A} = [a_{i,g}] \in \mathbb{R}^{N \times G}$, where $a_{i,g}$ is the expression of gene $g \in \{1, \ldots, G\}$ in spot/cell $i \in \{1, \ldots, N\}$,
- an $N \times 2$ spatial location matrix $\mathbf{S} = [\mathbf{s}_i] \in \mathbb{R}^{N \times 2}$, where $\mathbf{s}_i = (x_i, y_i)$ is the 2-D coordinate of spot/cell i for $i = 1, \ldots, N$,

and a list $\mathcal{L} \subseteq G \times G$ of putatively interacting genes; e.g. from a database of experimentally observed ligand-receptor interactions such as CellChatDB [27].

The ST data $(\mathbf{A}, \mathbf{S})$ is the standard output from applying a spatial sequencing or imaging technology to a 2-D tissue slice. Such spatial technologies include 10x Genomics Visium [28], Slide-Seq/Slide-SeqV2/Slide-tags [29, 8, 30], and MERFISH [31]. Different spatial technologies differ in the numbers N, G of spots and genes, respectively, that are measured, as well as their spatial resolution. See [32, 33, 34] for details. We also note that the list $\mathcal{L}$ of putatively interacting genes is incomplete, i.e. not all interacting ligands and receptors are included in $\mathcal{L}$, and may also be biased towards genes that are well-studied (i.e. ascertainment bias).

Steps of GLACIER. GLACIER consists of the following steps. First, GLACIER uses GASTON [16] to compute the isodepth d_i for each cell $i = 1, \ldots, N$ and forms a spatial DAG G, as described in Section 2.2 (Figure 1B, left). For each pair $(g, h) \in \mathcal{L}$ of candidate interacting genes, GLACIER uses the spatial DAG to test and report whether the observed gene-g expression values $\mathbf{a}_g = (a_{i,g})_{i=1}^N$ *Granger causes* the observed gene-h expression values $\mathbf{a}_h = (a_{i,h})_{i=1}^N$ (Figure 1B, right), and reports all pairs (g, h) where this Granger causality relationship holds (Figure 1C).

We test for Granger causality using a version of Velorama [17] adapted to spatial data. We construct the spatial DAG using a k-NN graph (with $k = 7$) on each cell's spatial coordinates, oriented using isodepth (Section 2.2). Notably, isodepth has the unique property that it can be *inverted* by redefining each cell's isodepth as

$$d_i' = \max_{1 \le j \le N} (d_j) - d_i \tag{6}$$

which reverses the ordering of isodepth among cells, allowing for analysis of information flow in the opposite direction.

Hyperparameter selection. Due to the sparsity of transcript counts from current ST technologies, several hyperparameter adjustments are necessary to ensure Velorama is able to run smoothly. We adjusted key hyperparameters as follows:

- **Learning Rate**: A higher learning rate than the default used in Velorama was sometimes necessary to achieve stable convergence. To balance this higher learning rate, we also implemented a learning rate scheduler.
- **Number of Epochs**: We also needed to increase the number of training epochs to allow the model sufficient time to reach an optimal solution due to the number of genes and learning rate scheduling.
- **Regularization (λ)**: Due to the sparsity of spatial transcriptomics data, we needed to sample from a broader range of λ values to ensure valid settings were obtained. As defined in [17], valid settings are those where between 5% and 95% of regularization weights remain nonzero.

To validate the robustness of our hyperparameter choices, we conducted a bootstrap-style analysis in which the dataset was randomly partitioned into three subsets. We selected hyperparameters that led to

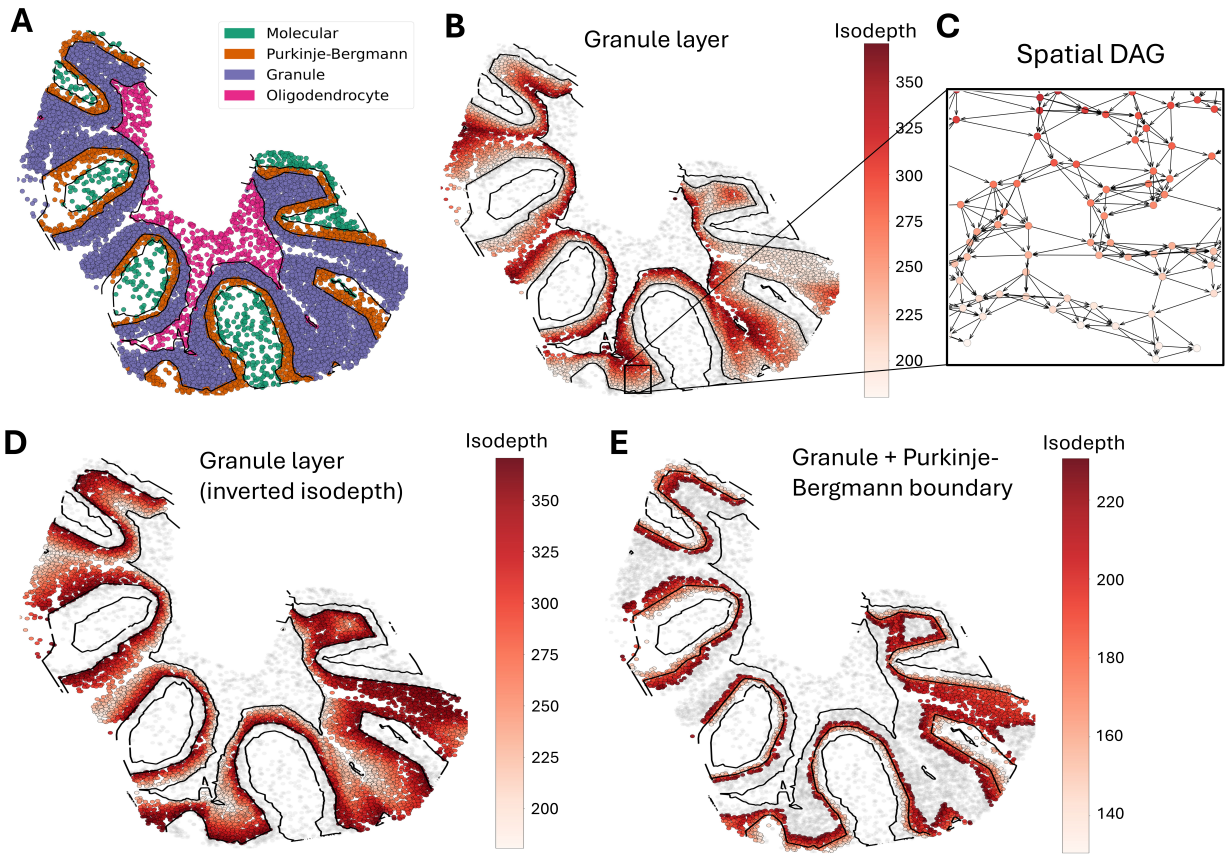

Figure 2: **(A)** Slide-SeqV2 mouse cerebellum dataset with each cell colored by its layer assigned by GASTON [16]. Each layer is named according to its dominant cell type as in [20, 35]. **(B)** Cells in granular layer colored by GASTON isodepth. Black curves denote GASTON-idenfied layer boundaries. **(C)** Spatial DAG $G = (V, E)$ constructed by GLACIER in granular layer, with directed edges (i, j) oriented from high isodepth to low isodepth, i.e. edge $i \longrightarrow j$ if $d_i > d_j$. **(D)** Inverted isodepth d_i' (equation (6)) in granular layer, which allows for identification of *bidirectional* cellular interactions by GLACIER. **(E)** Cells with isodepth d_i within $50\mu m$ of layer boundary between granular layer and Purkinje-Bergmann layer, colored by isodepth d_i, allowing for GLACIER to identify cellular interactions *between* cell types.

robust inference results across all three runs; see Appendix A for more details. We provide guidance on hyperparameter selection with the source code release. Running GLACIER on the full mouse cerebellum dataset, which consists of 9,985 cells, typically required ≈ 400 minutes on a single Nvidia A6000 GPU.

3 Results

We used GLACIER to identify cellular interactions in the mouse cerebellum using spatial transcriptomics (ST) data measured using the Slide-SeqV2 technology [8, 20]. The expression of 23, 096 transcripts was measured at 9, 985 spatial locations. The mouse cerebellum has a layered geometry and consists of four distinct layers – the oligodendrocyte, granular, Purkinje–Bergmann (PB) and molecular layers – as identified in previous studies [16, 20] (Figure 2A). To investigate transcription factor (TF)-target and ligand-receptor gene interactions, we selected 39 TFs and 773 target genes and 23 ligand genes and 26 receptor genes, respectively, from the GASTON spatially variable genes [16].

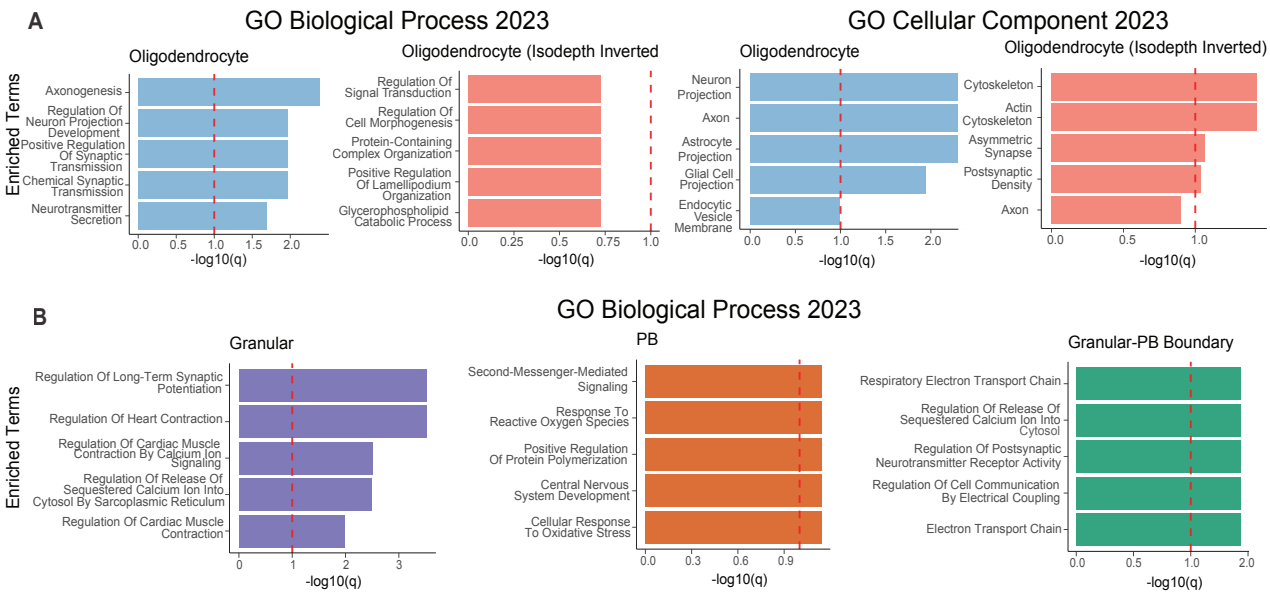

Figure 3: **(A)** Gene set enrichment analysis (GSEA) with Enrichr [36] for cellular interactions identified by GLACIER in the oligodendrocyte layer with isodepth d and inverted isodepth d' isodepth orientations for biological process (left) and cellular component (right) gene sets. **(B)** GSEA for cellular interactions identified by GLACIER in the granular layer (left), Purkinje-Bergmann (PB) layer (middle), and the boundary between the granular and PB layers (right).

We used GLACIER to focus on two biological questions, which were previously out of reach for existing methods studying cellular interactions:

1. What is the *directionality* of cellular interactions – and more generally, information flow – within a layer?

2. Is there a *discontinuity* in information flow across layers?

To answer the first question, we used GLACIER to identify cellular interactions *within* each cerebellum layer (e.g. granule layer, Figure 2B), which involves forming a spatial DAG (Figure 2C) oriented according to the GASTON isodepth d which smoothly varies within each layer. Importantly, by *inverting* the isodepth (Methods, Figure 2D), GLACIER identifies cellular interactions in the opposite direction and thus determine the *directionality* of cellular interactions. For the second question, we used GLACIER to identify cellular interactions that are unique to cells at the *boundary* of a cerebellum layer (Figure 2E).

To address these questions, we analyzed the data at multiple levels. We identified frequently-regulated target genes and performed gene set enrichment analysis (GSEA) to determine which biological processes were enriched among these targets [36]. We further examined specific TF-target and ligand-receptor gene pairs to explore key regulatory interactions at fine resolution. Lastly, we also exploited GLACIER-derived lag estimates to identify ligands with short- and long-ranging activity at layer boundaries.

3.1 Gene sets

Direction of information flow in oligodendrocytes: We examined gene set enrichments for oligodendrocytes in both orientations of isodepth. We observed that increasing isodepth, i.e. where the DAG leads away from the granular layer, exhibited a much more coherent and interpretable set of enriched biological

processes. This direction was strongly associated with axonogenesis, synaptic transmission, and neuronal connectivity, suggesting that oligodendrocytes in this orientation actively shape neuronal function rather than simply responding to their environment (Figure 3A-B). Function enrichments included axon development, regulation of neurotransmitter secretion, and neuron projection morphogenesis, all of which align with known roles of oligodendrocytes in supporting and refining neural circuits [37].

At the gene level, this enrichment was driven by regulators of axon guidance and synaptic adhesion, such as neurexin-1 (*Nrxn1*) and neurofascin (*Nfasc*), which mediate neuron-glia communication, and *Mag*, which plays a key role in stabilizing axonal connections. In contrast, the decreasing isodepth (inverted) direction showed weaker and less functionally coherent enrichments, with a shift toward intracellular and cytoskeletal remodeling rather than neuron-facing processes. Genes such as *Cfl1* and *Phip*, which regulate actin dynamics [38] and intracellular signaling, were more prominent in this direction, suggesting a greater emphasis on structural maintenance or adaptation.

These findings raise the possibility that oligodendrocytes' instructive role proceeds in the direction of isodepth, with less regulatory control exerted in the other direction. Further experimental work would be needed to investigate this hypothesis. Indeed, the value of GLACIER is precisely in helping generate such hypotheses for further investigation.

Regulatory discontinuity at the boundary of granular and Purkinje-Bergmann layers: Gene set enrichment at this boundary revealed a distinct functional signature that differs from either layer alone. While the granular layer was enriched for processes related to long-term synaptic potentiation and calcium-mediated excitability (Figure 3C, left), and the PB layer showed enrichments for oxidative stress response and second-messenger signaling (Figure 3C, middle), the boundary exhibited a distinct focus on synaptic modulation, electrical coupling, and metabolic activity (Figure 3C, right). This suggests that cells at the interface of the granular and Purkinje-Bergmann layers play a specialized role in mediating communication and bioenergetic demands between the two layers.

One of the strongest enrichments at the boundary included regulation of postsynaptic neurotransmitter receptor activity and cell communication by electrical coupling, suggesting rapid synaptic transmission. We also observed enrichment of electron transport chain activity possibly due to metabolic interactions. The gene set included *Dlgap1* and *Nptx1*, both of which are involved in synaptic scaffolding and activity-dependent synapse formation, reinforcing the idea that the boundary plays a role in synaptic refinement and communication [39]. In addition, *Ndufa5* and *Sdha*, key components of the mitochondrial respiratory chain [40], indicate that cells in this region experience high metabolic demand and are likely to support sustained neurotransmission.

These results suggest that the Purkinje-Bergmann and granular cell boundary is not simply a gradient between two transcriptional programs, but rather a specialized interface with its own regulatory demands. Experimental validation could determine whether these transcriptional differences are critical to sustaining high-frequency signaling and synaptic communication in the cerebellum.

3.2 Gene pairs

GLACIER reveals several biologically meaningful pairs of genes (TF-target or ligand-receptor) with spatial Granger causal relationships within individual layers. For example, GLACIER identifies that the TF *Hspa5* is spatially Granger causal for the target gene *Pink1* (Figure 4A), with *Pink1* expression closely following *Hspa5* expression. *Hspa5* maintains endoplasmic reticulum (ER) homeostasis [41] while *Pink1* performs mitochondrial quality control [42], suggesting that activation of general ER regulation results in recruitment of proteins that regulate specific cellular functions. GLACIER also identifies that the ligand *S100b* is spatially Granger causal of the receptor *Ptprz1* (Figure 4B). *S100b* has been suggested to support neuron migration [43] while *Ptprz1* has been observed to modulate cell migration in some tissues [44], suggesting

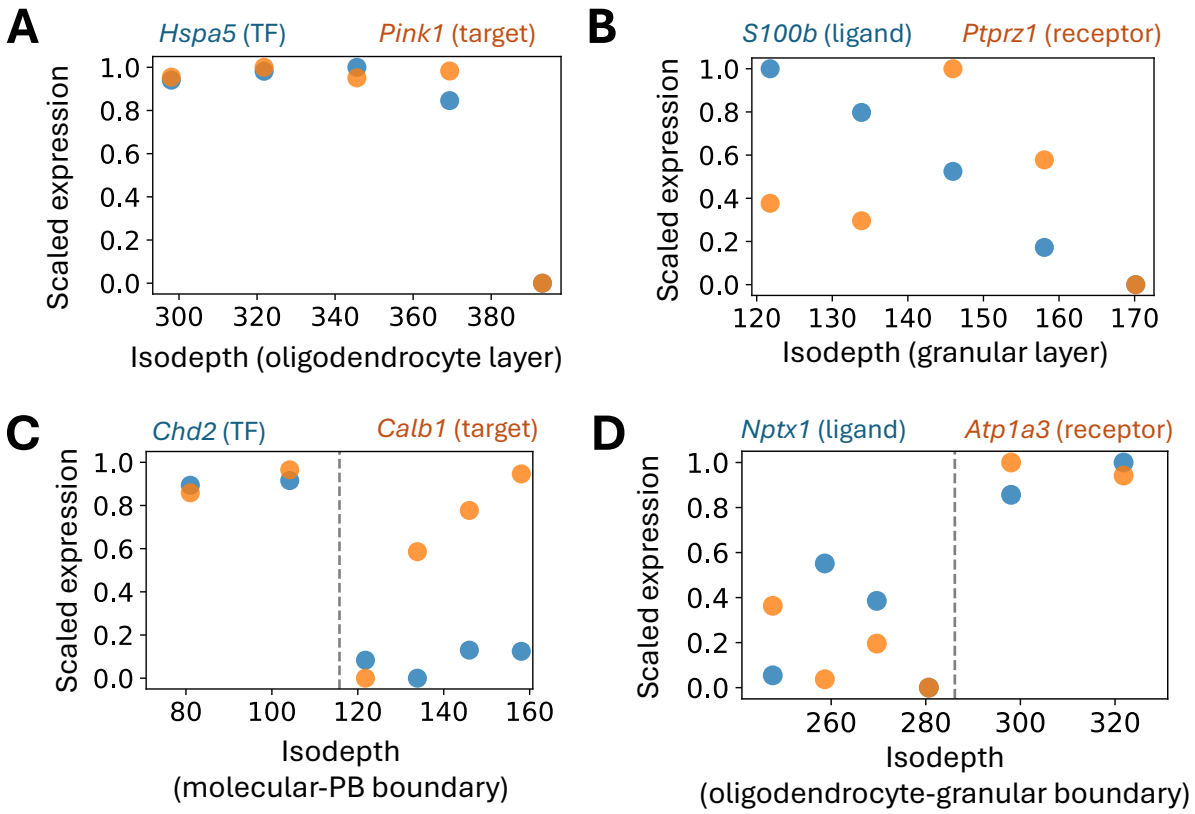

Figure 4: Isodepth versus scaled expression for **(A)** transcription factor *Hspa5* and target gene *Pink1* in the oligodendrocyte layer, **(B)** ligand *S100b* and target gene *Ptprz1* in the granular layer, **(C)** transcription factor *Chd2* and target gene *Calb1* at the boundary between the molecular and Purkinje-Bergmann (PB) layers, and **(D)** ligand *Nptx1* and receptor *Atp1a3* at the boundary between the oligodendrocyte and granular layers.

that neurons in the granular layer may be moving in the isodepth direction towards the adjacent oligodendrocyte layer.

Further, GLACIER identifies spatially Granger causal pairs of genes that are unique to the *boundary* between two layers. For example, GLACIER identifies an interaction between the TF *Chd2* and target gene *Calb1* at the boundary between the molecular and Purkinje-Bergmann (PB) layers (Figure 4C). *Chd2* has been observed to broadly regulate neuronal function [45] while *Calb1* is a marker gene for Purkinje cells [46], suggesting that *Chd2* may be regulating the function of Purkinje cells specifically at the molecular-PB layer boundary. GLACIER also identifies an interaction between the ligand *Nptx1* and the receptor *Atp1a3* that uniquely occurs at the boundary of the oligodendrocyte and granular layers; both *Nptx1* and *Atp1a3* are known to be involved in neuronal signaling and excitation [47, 48], suggesting potential communication between oligodendrocytes and granular cells at the boundary.

In this way, GLACIER identifies cellular interactions within and across layer boundaries which cannot be found by existing approaches which do not incorporate the global tissue topography.

3.3 Spatial range estimation via GLACIER lag estimates

GLACIER's causal inference framework reports not only the regulatory interactions but also the lag at which the interaction operates (Methods). This allows us to estimate the spatial range of a regulator gene (i.e. a TF or ligand), with bigger lags corresponding to a larger spatial range of a regulator. We investigated some settings where GLACIER reported bimodal lags across TF or ligand regulators (Figure S1), denoting substantial variations in the spatial ranges of regulators.

Molecular Layer/Purkinje-Bergmann Boundary: The boundary between the Molecular and Purkinje-Bergmann layers displayed particularly pronounced bimodality in the ligand distance distribution. The longest-ranging ligand at this interface was *Apoe*, a gene well-known for its role in lipid transport, synaptic plasticity, and neuronal maintenance [49] *Apoe*'s ability to bind lipoproteins and facilitate cholesterol uptake is crucial for neuronal repair and may underlie its long-range effects across these layers. Conversely, the shortest-acting ligand we identified was *Gdf10*, a member of the TGF-β superfamily involved in neuronal regeneration and repair [50]. *Gdf10*'s localized expression and role in modulating cell growth suggest its influence is confined to more proximal targets in neighboring cells.

Purkinje-Bergmann/Granular Layer Boundary: We also observed substantial spatial-range variation between ligands at the Purkinje-Bergmann/Granular layer boundary. Here, the furthest-acting ligand was *Gng13*, a *G*-protein gamma subunit that partners with alpha and beta subunits to regulate a variety of downstream signaling pathways [51]. Its presence in long-range signaling contexts suggests that *Gng13* may orchestrate interactions spanning multiple cell types or spatial zones in the cerebellar architecture. While there was no single dominant short-range ligand at this boundary, the *Thy1* receptor was repeatedly seen. *Thy1* (also known as *Cd90*) is a cell-surface glycoprotein involved in cell adhesion, neurite outgrowth, and T-cell activation [52]. Its role in modulating cell-cell contact may explain its consistent appearance among ligands operating over limited spatial distances, hinting that *Thy1*-dependent interactions mediate localized communication in these cerebellar layers.

Taken together, these findings underscore how spatial organization can drive bimodal patterns of ligand-receptor signaling: some factors, like *Apoe* and *Gng13*, can reach distant targets, while others, such as *Gdf10*, *Gnas*, and *Thy1*, function in tight, localized niches. By systematically exploring these "lag distributions," GLACIER reveals how cerebellar architecture shapes the breadth and specificity of cellular communication.

4 Discussion

GLACIER introduces a framework for inferring causal transcriptional and signaling relationships in spatial transcriptomics by leveraging a global tissue topology. Unlike previous approaches that rely on local cell-cell adjacency, GLACIER constructs a directed acyclic graph (DAG) based on a global spatial coordinate, allowing us to infer how regulatory programs unfold along natural tissue axes. By orienting regulatory interactions along inferred spatial axes, we move beyond local spatial correlations and instead model how gene regulation and cell signaling propagate through *global* tissue structure. Applied to the cerebellum, GLACIER uncovered directional transcriptional programs within layers and distinct regulatory activity at layer boundaries, demonstrating its ability to identify spatially organized cellular interactions.

One limitation of GLACIER is that it requires tissues where a global spatial coordinate, such as isodepth, can be inferred. However, this is not a fundamental restriction – user-guided segmentation of complex tissues can enable locally defined spatial coordinates, allowing GLACIER to be applied more broadly. Moreover, one can also use a supervised approach for determining such spatial coordinates, such as in [53, 54]. Another challenge is transcript sparsity, a common issue in spatial transcriptomics that may become more pronounced with technologies like Xenium [55]. However, GLACIER naturally mitigates this issue by

aggregating over multiple cells (i.e. cells with same isodepth), increasing robustness to sparsity. Additionally, GLACIER's bootstrap-based hyperparameter selection provides a systematic way to ensure robust inferences.

While we applied GLACIER's spatial Granger causality approach to brain layers, the methodology generalizes to other structured tissues. Future work could leverage this framework to study how signaling cascades change radially from blood vessels or how the interior and periphery of a tumor microenvironment exhibit distinct regulatory states. Looking forward, integrating GLACIER with spatial proteomics [56] and multimodal single-cell datasets [57] could provide a more comprehensive view of spatially structured regulatory programs, further enhancing our ability to decode the organization of complex tissues.

Acknowledgments

U.C. was supported in part by funding from the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard. H.L. and R.S. were partially supported by the Chan Zuckerberg Initiative.

References

[1] Jimeng Su, Ying Song, Zhipeng Zhu, Xinyue Huang, Jibiao Fan, Jie Qiao, and Fengbiao Mao. Cell–cell communication: new insights and clinical implications. *Signal Transduction and Targeted Therapy*, 9(1):196, 2024.

[2] Urs Mayr, Denise Serra, Prisca Liberali, Allon Klein, and Barbara Treutlein. Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development*, 146(12), 2019.

[3] Christine Yiwen Yeh, Karmen Aguirre, Olivia Laveroni, Subin Kim, Aihui Wang, Brooke Liang, Xiaoming Zhang, Lucy M Han, Raeline Valbuena, Michael C Bassik, et al. Mapping spatial organization and genetic cell-state regulators to target immune evasion in ovarian cancer. *Nature immunology*, 25(10):1943–1958, 2024.

[4] Zhuang Liu, Xiaowei Bian, Lihua Luo, Åsa K Björklund, Li Li, Letian Zhang, Yongjian Chen, Lei Guo, Juan Gao, Chunyan Cao, et al. Spatiotemporal single-cell roadmap of human skin wound healing. *Cell Stem Cell*, 2024.

[5] Sabrina M Lewis, Marie-Liesse Asselin-Labat, Quan Nguyen, Jean Berthelet, Xiao Tan, Verena C Wimmer, Delphine Merino, Kelly L Rogers, and Shalin H Naik. Spatial omics and multiplexed imaging to explore cancer biology. *Nature methods*, 18(9):997–1012, 2021.

[6] Hirak Sarkar, Uthsav Chitra, Julian Gold, and Benjamin J Raphael. A count-based model for delineating cell–cell interactions in spatial transcriptomics data. *Bioinformatics*, 40(Supplement_1):i481–i489, 2024.

[7] Mihir Bafna, Hechen Li, and Xiuwei Zhang. Clarify: cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics*, 39(Supplement_1):i484–i493, 2023.

[8] Robert R Stickels, Evan Murray, Pawan Kumar, Jilong Li, Jamie L Marshall, Daniela J Di Bella, Paola Arlotta, Evan Z Macosko, and Fei Chen. Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. *Nature biotechnology*, 39(3):313–319, 2021.

[9] Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

[10] Ruben Dries, Qian Zhu, Rui Dong, Chee-Huat Linus Eng, Huipeng Li, Kan Liu, Yuntian Fu, Tianxiao Zhao, Arpan Sarkar, Feng Bao, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology*, 22:1–31, 2021.

[11] Brendan F Miller, Dhananjay Bambah-Mukku, Catherine Dulac, Xiaowei Zhuang, and Jean Fan. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome research*, 31(10):1843–1855, 2021.

[12] Duy Pham, Xiao Tan, Brad Balderson, Jun Xu, Laura F Grice, Sohye Yoon, Emily F Willis, Minh Tran, Pui Yeng Lam, Arti Raghubar, et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature communications*, 14(1):7739, 2023.

[13] Zhuoxuan Li, Tianjie Wang, Pentao Liu, and Yuanhua Huang. Spatialdm for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nature communications*, 14(1):3995, 2023.

[14] Chong Wang, Brendon M Baker, Christopher S Chen, and Martin Alexander Schwartz. Endothelial cell sensing of flow direction. *Arteriosclerosis, thrombosis, and vascular biology*, 33(9):2130–2136, 2013.

[15] Eric C Lai. Notch signaling: control of cell communication and cell fate. 2004.

[16] Uthsav Chitra, Brian J. Arnold, Hirak Sarkar, Kohei Sanno, Cong Ma, Sereno Lopez-Darwin, and Benjamin J. Raphael. Mapping the topography of spatial gene expression with interpretable deep learning. *Nature Methods*, 2025.

[17] Rohit Singh, Alexander P Wu, Anish Mudide, and Bonnie Berger. Causal gene regulatory analysis with rna velocity reveals an interplay between slow and fast transcription factors. *Cell Systems*, 15(5):462–474, 2024.

[18] Alexander P Wu, Rohit Singh, and Bonnie Berger. Granger causal inference on DAGs identifies genomic loci regulating transcription. In *International Conference on Learning Representations*, 2022.

[19] Alexander Po-Yen Wu, Rohit Singh, Christopher A Walsh, and Bonnie Berger. An econometric lens resolves cell-state parallax. *bioRxiv*, pages 2023–03, 2023.

[20] Dylan M Cable, Evan Murray, Luli S Zou, Aleksandrina Goeva, Evan Z Macosko, Fei Chen, and Rafael A Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature biotechnology*, 40(4):517–526, 2022.

[21] Walter Enders. *Applied econometric time series.* John Wiley & Sons, 2008.

[22] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[23] Michael W McCracken. Asymptotics for out of sample tests of granger causality. *Journal of econometrics*, 140(2):719–752, 2007.

[24] Ali Shojaie and Emily B Fox. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9(1):289–319, 2022.

[25] Zhenqin Wu, Alexandro E Trevino, Eric Wu, Kyle Swanson, Honesty J Kim, H Blaize D'Angio, Ryan Preska, Gregory W Charville, Piero D Dalerba, Ann Marie Egloff, et al. Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nature Biomedical Engineering*, 6(12):1435–1448, 2022.

[26] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B. Fox. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2022.

[27] Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, and Qing Nie. Inference and analysis of cell-cell communication using cellchat. *Nature communications*, 12(1):1088, 2021.

[28] 10x Visium Genomics Visium Spatial Gene Expression. https://www.10xgenomics.com/products/spatial-gene-expression.

[29] Samuel G Rodriques, Robert R Stickels, Aleksandrina Goeva, Carly A Martin, Evan Murray, Charles R Vanderburg, Joshua Welch, Linlin M Chen, Fei Chen, and Evan Z Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

[30] Andrew JC Russell, Jackson A Weir, Naeem M Nadaf, Matthew Shabet, Vipin Kumar, Sandeep Kambhampati, Ruth Raichur, Giovanni J Marrero, Sophia Liu, Karol S Balderrama, et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature*, pages 1–9, 2023.

[31] Jeffrey R Moffitt, Dhananjay Bambah-Mukku, Stephen W Eichhorn, Eric Vaughn, Karthik Shekhar, Julio D Perez, Nimrod D Rubinstein, Junjie Hao, Aviv Regev, Catherine Dulac, et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, 362(6416):eaau5324, 2018.

[32] Luyi Tian, Fei Chen, and Evan Z Macosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782, 2023.

[33] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

[34] Giovanni Palla, David S Fischer, Aviv Regev, and Fabian J Theis. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, 2022.

[35] Dylan M Cable, Evan Murray, Vignesh Shanmugam, Simon Zhang, Luli S Zou, Michael Diao, Haiqi Chen, Evan Z Macosko, Rafael A Irizarry, and Fei Chen. Cell type-specific inference of differential expression in spatial transcriptomics. *Nature methods*, 19(9):1076–1087, 2022.

[36] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.

[37] Mikael Simons and Klaus-Armin Nave. Oligodendrocytes: Myelination and axonal support. *Cold Spring Harbor Perspectives in Biology*, 8(1):a020479, 2015.

[38] Diana C Muñoz-Lasso, Carlos Romá-Mateo, Federico V Pallardó, and Pilar Gonzalez-Cabo. Much more than a scaffold: cytoskeletal proteins in neurological disorders. *Cells*, 9(2):358, 2020.

[39] Andreas H Rasmussen, Hanne B Rasmussen, and Asli Silahtaroglu. The dlgap family: neuronal expression, function and role in brain disorders. *Molecular Brain*, 10:43, 2017.

[40] Alessandra Torraco, Susana Peralta, Luisa Iommarini, and Francisca Diaz. Mitochondrial diseases part i: mouse models of oxphos deficiencies caused by defects in respiratory complex subunits or assembly factors. *Mitochondrion*, 21:76–91, 2015.

[41] Jie Wang, Jessica Lee, David Liem, and Peipei Ping. Hspa5 gene encoding hsp70 chaperone bip in the endoplasmic reticulum. *Gene*, 618:14–23, 2017.

[42] Alicia M Pickrell and Richard J Youle. The roles of pink1, parkin, and mitochondrial fidelity in parkinson's disease. *Neuron*, 85(2):257–273, 2015.

[43] Björn Bluhm, Björn Laffer, Daniela Hirnet, Matthias Rothermundt, Oliver Ambree, and Christian Lohr. Normal cerebellar development in s100b-deficient mice. *The Cerebellum*, 14:119–127, 2015.

[44] Kan V Lu, Kimberly A Jong, Gloria Y Kim, Jatinder Singh, Ederlyn Q Dia, Koji Yoshimoto, Maria Y Wang, Timothy F Cloughesy, Stanley F Nelson, and Paul S Mischel. Differential induction of glioblastoma migration and growth by two forms of pleiotrophin. *Journal of Biological Chemistry*, 280(29):26953–26964, 2005.

[45] EMA Lewis, G Chapman, K Kaushik, J Determan, I Antony, K Meganathan, M Narasimhan, P Gontarz, B Zhang, and KL Kroll. Regulation of human cortical interneuron development by the chromatin remodeling protein chd2. *Scientific reports*, 12(1):15636, 2022.

[46] Pegah Afshar, Niloufar Ashtari, Xiaodan Jiao, Maryam Rahimi-Balaei, Xiaosha Zhang, Behzad Yaganeh, Marc R Del Bigio, Jiming Kong, and Hassan Marzban. Overexpression of human sod1 leads to discrete defects in the cerebellar architecture in the mouse. *Frontiers in neuroanatomy*, 11:22, 2017.

[47] Marie Coutelier, Maxime Jacoupy, Alexandre Janer, Flore Renaud, Nicolas Auger, Ganapathi-Varma Saripella, François Ancien, Fabrizio Pucci, Marianne Rooman, Dimitri Gilis, et al. Nptx1 mutations trigger endoplasmic reticulum stress and cause autosomal dominant cerebellar ataxia. *Brain*, 145(4):1519–1534, 2022.

[48] Erin L Heinzen, Alexis Arzimanoglou, Allison Brashear, Steven J Clapcote, Fiorella Gurrieri, David B Goldstein, Sigurdur H Johannesson, Mohamad A Mikati, Brian Neville, Sophie Nicole, et al. Distinct neurological disorders with atp1a3 mutations. *The Lancet Neurology*, 13(5):503–514, 2014.

[49] Robert W. Mahley. Apolipoprotein e: from cardiovascular disease to neurodegenerative disorders. *Journal of Molecular Medicine*, 94(7):739–746, 2016.

[50] S. Li, E. H. Nie, Y. Yin, L. I. Benowitz, S. Tung, H. V. Vinters, F. R. Bahjat, M. P. Stenzel-Poore, R. Kawaguchi, G. Coppola, and S. T. Carmichael. GDF10 is a signal for axonal sprouting and functional recovery after stroke. *Nature Neuroscience*, 18(12):1737–1745, 2015.

[51] David E. Clapham and Eliot J. Neer. G protein beta gamma subunits. *Annual Review of Pharmacology and Toxicology*, 37:167–203, 1997.

[52] Tanya A. Rege and James S. Hagood. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, and cancer progression. *FASEB Journal*, 20(8):1045–1054, 2006.

[53] Cong Ma, Uthsav Chitra, Shirley Zhang, and Benjamin J Raphael. Belayer: Modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Systems*, 13(10):786–797, 2022.

[54] Uthsav Chitra, Shu Dan, Fenna M Krienen, and Benjamin J Raphael. Gaston-mix: a unified model of spatial gradients and domains using spatial mixture-of-experts. *bioRxiv*, pages 2025–01, 2025.

[55] Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature Communications*, 14(1):8353, 2023.

[56] Emma Lundberg and Georg HH Borner. Spatial proteomics: a powerful discovery tool for cell biology. *Nature Reviews Molecular Cell Biology*, 20(5):285–302, 2019.

[57] Katy Vandereyken, Alejandro Sifrim, Bernard Thienpont, and Thierry Voet. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics*, 24(8):494–515, 2023.

Appendix

A Robustness testing

Velorama sweeps over a range of regularization hyperparameters (λ), averaging TF-target gene interactions from models where the percentage of non-zero interactions fall within a given threshold (by default, 0.01 to 0.95). To find λ values meeting this threshold for each layer and isodepth orientation, we increased the number of λs tested from the default of 10 to 30. We also narrowed the search range from $(0.01, 10)$ to $(0.001, 0.1)$, since we found suitable λs was always below 0.1 empirically. We used a higher learning rate ($\eta = 0.1$) for TF-target gene analysis and a lower learning rate ($\eta = 0.01$) for ligand-receptor analysis. The changes in learning rate were made across all runs, and the number of iterations per λ during training was significantly increased from the default of 1000 to between 10,000 and 30,000, depending on the run.

We verified other hyperparameters by testing on stratified thirds of the dataset. Each stratified third was constructed by evenly selecting cells from cell type-specific subsets and then combining them, ensuring that the cell type distribution was preserved. We tuned hyperparameters so that results were consistent across stratified thirds and the overall dataset (for both normal and inverted isodepth). Finally, we analyzed the loss for each run to confirm smooth training behavior and convergence at the end of each model training run.

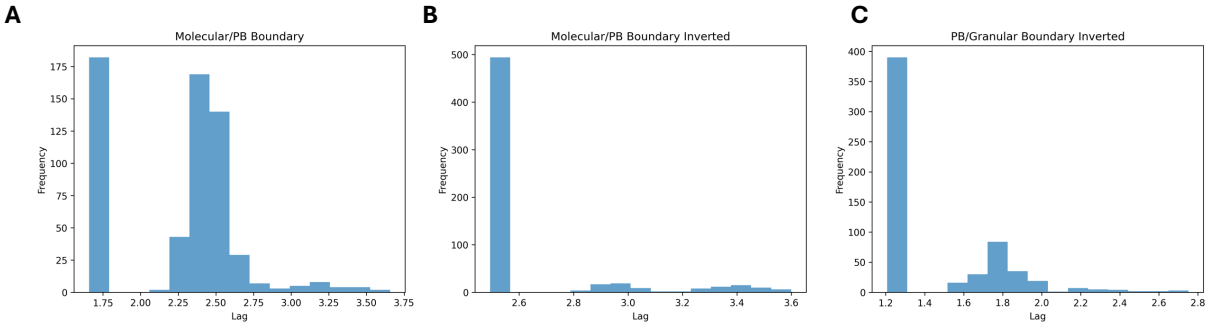

Figure S1: Distribution of lags identified by GLACIER across all candidate interactions in **(A)** the boundary between the molecular and Purkinje-Bergmann (PB) layers using the isodepth d, **(B)** the boundary between the molecular and PB layers using the inverted isodepth d', and **(C)** the boundary between the PB and granular layers using the inverted isodepth d'.